DOCUMENT RESUME

ED 422 802 HE 031 560

AUTHOR Underwood, David G.; Craighead, Michelle M.
TITLE Graduate Record Examination (GRE) Scores as an Assessment Tool. AIR 1998 Annual Forum Paper.
PUB DATE 1998-05-00
NOTE 15p.; Paper presented at the Annual Forum of the Association for Institutional Research (38th, Minneapolis, MN, May 17-20, 1998).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Accountability; *Accreditation (Institutions); *College Outcomes Assessment; Correlation; Evaluation Criteria; Higher Education; *Institutional Evaluation; Majors (Students); Outcome Based Education; Performance; Predictive Measurement; Regression (Statistics); *School Effectiveness
IDENTIFIERS *AIR Forum; *Graduate Record Examinations; Scholastic Assessment Tests

ABSTRACT

This paper examines whether graduate record examination (GRE) scores are a legitimate assessment tool for measuring institutional accountability and effectiveness that is, how well its graduates will do after having attended the institution for four or more years. Following a discussion of the various pros and cons of using the GRE as an accountability measure, the study reports on an examination of all GRE reports for five years (May 1992-May 1997) collected at a land-grant research university in the southeast (n=2,934). Regression models were developed using Scholastic Assessment Test verbal and math scores, gender, race, cumulative credit hours, and grade point averages to create predicted GRE total, quantitative, verbal, and analytical scores. Then the predicted GRE scores were subtracted from the actual GRE scores to provide a residual score, which was analyzed by major to determine whether any of the residuals were greater than expected through random variation. Significant differences were found to exist based on the mean of the residuals by major, and these were further analyzed. The report concludes that using this assessment approach leaves unanswered the question of whether the information garnered can be used to improve programs and services of the institution. (Contains 7 references.)(CH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 422 802

Graduate Record Examination (GRE) Scores as an Assessment Tool

David G. Underwood

Director of Assessment

B-17 Hardin Hall, Box 345155

Clemson University

Clemson, South Carolina 29634-5155

(864) 656-0868

Michelle M. Craighead

Research Graduate Assistant

B-17 Hardin Hall, Box 345155

Clemson University

Clemson, South Carolina 29634-5155

(864) 656-1410

IR 031 560

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

AIR

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

for Management Research, Policy Analysis, and Planning

This paper was presented at the Thirty-Eighth Annual Forum of the Association for Institutional Research held in Minneapolis, Minnesota, May 17-20, 1998.
This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of AIR Forum Papers.

Dolores Vura
Editor
AIR Forum Publications

Graduate Record Examination (GRE) Scores as an Assessment Tool

Abstract

Each year this research II, land-grant university subscribes to a service of Educational Testing Service (ETS) to receive regular updates of Graduate Record Examination (GRE) scores from individuals who wish to attend, or have previously graduated from, this institution. In addition to the subscription expense, there is an additional expense involved in getting the results into a database and maintaining it. This paper focuses on making a determination of whether the GRE scores are a legitimate assessment tool. A discussion of GRE scores is provided along with the results of a study to determine whether meaningful information can be provided by a "talent development perspective" suggested by Alexander Astin.

Introduction

For more than a decade, postsecondary institutions have been faced with increasing demands for accountability. These demands come both from regional accrediting bodies such as the Southern Association of Colleges and Schools (SACS) and mandates from state legislatures. With SACS the accountability function falls under the auspices of "institutional effectiveness," with an emphasis on use of the results for continued improvement of programs and services. As a result of the increased focus on accountability, most states have mandated some type of assessment activities requiring institutions to demonstrate accountability for the graduates they produce. Quite often the focus of the legislated accountability is on the reporting of numbers rather than the improvement of programs and services.

Although SACS, as well as most other regional accrediting bodies, does not specify what data should be collected to demonstrate institutional effectiveness, they do provide a list of types of data which could be used. Because the data collection methods and the types of data to be collected are not specified, institutions have struggled with decisions about which types of data to collect, how to collect them, and when to collect them. In the cases of legislated accountability, the methods and the data types are often clearly specified. In attempts to help clarify what institutions might use to provide evidence of accountability, several authors have compiled lists, or checklists, of data types that might be used (Bottrill & Borden, 1994; Jacobi, Astin, & Ayala, 1987; Nichols, 1991). Nearly all of these lists suggest the Graduate Record Examination (GRE) scores as an indicator which institutions could use. The legislature in South Dakota went so far as to mandate the reporting of GRE scores as part of institutional accountability in that state (Banta, 1993).

Why the popularity of GRE scores as an indicator of either accountability or institutional effectiveness? Availability is probably one of the top reasons. Also, since the GRE is completed after the college program, the common belief appears to be that if the institution provided a "quality" education, then individuals who take the GRE will have that reflected in their scores. Since it is a nationally normed, standardized examination, it is relatively easy to determine how the graduates from a particular institution compare with others. However, for an outcome indicator to be useful for assessment purposes, it should meet several criteria: 1) it must be accessible to the institution with relatively few resource costs, 2) it should provide some unique insights into the programs or processes of the institution above and beyond other information which is already available, and 3) it should provide information which is detailed enough to allow the institution to make changes to improve programs.

The GRE scores are accessible. For a relatively small fee, Educational Testing Service which produces the GRE, will provide an institution with score reports, both on paper and in a data file, which can then be used for additional analyses. At this institution, and the authors suspect at many others, the additional analyses consist of providing mean scores broken out by college or department. Those mean scores are then compared to national norms to determine how well the institution is doing in preparing graduates.

Whether the GRE provides unique insights not available through other sources is a more difficult question. Several studies point to the fact that GRE scores and Scholastic Aptitude Test (SAT) scores are very highly correlated (Angoff & Johnson, 1988; Astin, 1991). In the study by Angoff and Johnson (1988), the correlation between the two sets of scores was reported to be .86, indicating that approximately 74% of the variation in GRE scores could be

accounted for by knowledge of the SAT scores. Simply translated, this means that how well a graduate will do on the GRE, after having been affected by the institution for four or more years, can be very accurately predicted by knowing how well the individual did on the SAT examination prior to enrolling in college. Another way of thinking about this is that the institution can not, without further analysis, take credit for producing a given score on the GRE since the majority of that score appears to be based on entering characteristics of the individual and not on anything the institution provided. This finding was also supported by Alexander and Stark (1986), "Apparently, student characteristics are more predictive of GRE area scores than institutional characteristics. This finding indicates that changes in learning may not be attributed to institutional characteristics, but perhaps must be examined at a lower programmatic level," (p. 18). Studies have also found that the score is related to the gender of the test-taker with males scoring significantly higher on the quantitative portion than females (Angoff & Johnson, 1988).

The GRE provides three scores to the institution: a verbal score, a quantitative score and an analytical score. These three are often combined to form an additional score representing the total (Verbal + Quantitative + Analytical). The meaningfulness of these scores for curriculum or program decisions is highly questionable. Since the scores are not broken down into specific areas within each category, a problem pointed out several years ago by Jacobi, Astin, and Ayala (1987), the institution is provided with little useful information. For example, if the scores on the quantitative area are not as high as the institution would hope, there is no way to determine what must be enhanced in the curriculum. The low scores could come from a weakness in basic math, algebra, trigonometry, etc., but this level of information is not provided. Thus, the scores do little to provide useful information at the curriculum level.

Since the GRE appears to add little unique information, and since it can not be broken down to the curriculum level, its utility as an assessment tool becomes questionable. However, some studies have found that the GRE score is related to course taking patterns (Angoff and Johnson, 1988) and that aspect may lend itself to an assessment approach. "The impact of curriculum and sex was found to be low on GRE-verbal scores, but relatively high for GRE-quantitative, with students in highly quantitative fields enjoying an advantage over their peers in less quantitative fields of study." (p. i).

The course taking pattern is related to the idea of Jacobi, Astin, and Ayala's "talent development perspective" (1987) in which the focus is not on how well a student scores on an examination, but rather on the difference between what the student scores and what he or she was "expected" to score. Alexander Astin (1991) discusses this perspective in much more detail and makes specific recommendations about how to use the GRE in a "talent development" approach to assessment. The current study uses Astin's approach to determine whether the GRE scores can be useful with additional analyses.

Methodology

All of the GRE score reports for five years (May 1992 through May 1997) of a research II, land-grant university in the Southeast were collected resulting in 5,381 unduplicated scores. These scores were then matched with enrollment data from the student database using the social security number of the test-taker. This process allowed the extraction of grade point average (GPA), major field of study, the entering scores on the each area of the SAT, race, cumulative credit hours earned at graduation and gender. A total of 2,934 useable scores were obtained after matching.

As a first step in the analysis, following Astin's (1991) recommendations and based on his previous findings as well as those of Angoff and Johnson (1988), regression models were developed using the SAT verbal, and SAT math scores, gender, race, cumulative credit hours and grade point average of the individuals to create predicted GRE total, quantitative, verbal and analytical scores. The analyses were conducted using the Statistical Analysis System (SAS) regression procedure with a stepwise selection model.

In the second step, the predicted GRE was subtracted from the actual GRE score to provide a difference score, or residual. The residuals were then analyzed by major using a SAS Means procedure to determine whether any of the residuals were greater than expected through random variation. Then, the SAS General Linear Model (GLM) procedure was used to determine whether any of the differences between the mean expected scores and the mean actual scores were significantly different as a result of the student's major. In each case, the procedure identified significant differences between majors. As a result, post hoc analyses using Tukey's Honestly Significant Difference (HSD) were conducted to identify which majors were significantly different.

Findings

The model accounted for 71% of the variance in the GRE Total score, 69% of the variance in the GRE Quantitative score, 66.1% of the variance in the GRE Verbal score and 43.5% of the variance in the GRE Analytical score. See Tables 1 through 4 for model details.

Table 1

Summary of Stepwise Regression Analyses for Variables Predicting the GRE Total Score

(N=2,537)

Variable	β	Partial R^2	Model R^2
SAT Math	0.498	0.584	0.584
SAT Verbal	0.378	0.111	0.695
Gender (Female)	-0.075	0.005	0.700
GPA	0.079	0.005	0.705
Race (Minority)	-0.039	0.003	0.708
Age	0.041	0.002	0.710

Table 2

Summary of Stepwise Regression Analyses for Variables Predicting the GRE Quantitative Score (N=2,537)

Variable	β	Partial R^2	Model R^2
SAT Math	0.691	0.653	0.653
Gender (Female)	-0.173	0.024	0.677
GPA	0.096	0.008	0.685
CUMCREDIT	0.043	0.002	0.687
Age	0.035	0.001	0.688
SAT Verbal	0.039	0.001	0.689
Race (Minority)	-0.026	0.001	0.690

Table 3

Summary of Stepwise Regression Analyses for Variables Predicting the GRE Verbal Score

(N=2,537)

Variable	β	Partial R^2	Model R^2
SAT Verbal	0.745	0.639	0.639
Age	0.100	0.010	0.649
Gender (Female)	-0.064	0.005	0.654
GPA	0.067	0.005	0.659
SAT Math	0.041	0.001	0.660
Race (Minority)	-0.029	0.001	0.661

Table 4

Summary of Stepwise Regression Analyses for Variables Predicting the GRE Analytical Score

(N=2,537)

Variable	β	Partial R^2	Model R^2
SAT Math	0.450	0.365	0.365
SAT Verbal	0.257	0.056	0.421
Race (Minority)	-0.090	0.008	0.429
Sex (Female)	0.053	0.004	0.433
GPA	0.036	0.001	0.434
CUMCREDIT	-0.027	0.001	0.435

The mean residual, the mean of the differences between the actual and predicted scores for each group, was analyzed using the SAS means procedure. The means procedure was used to determine whether the differences were greater than would be expected due to random variation.

The residuals were also analyzed using the SAS GLM Procedure to conduct an analysis of variance using the major of the individual as a classification variable. The resulting ANOVA

provides an indication of whether the differences in the variation of the residuals might be a result of the individual's major. The results of the ANOVA appear in Tables 5 through 8.

Table 5

Summary of Analysis of Variance of GRE Total Scores using Undergraduate Major

Source	DF	F	PR > F
Undergraduate Major	13	5.06	0.0001

Table 6

Summary of Analysis of Variance of GRE Quantitative Scores using Undergraduate Major

Source	DF	F	PR > F
Undergraduate Major	13	13.68	0.0001

Table 7

Summary of Analysis of Variance of GRE Verbal Scores using Undergraduate Major

Source	DF	F	PR > F
Undergraduate Major	13	3.68	0.0001

Table 8

Summary of Analysis of Variance of GRE Analytical Scores using Undergraduate Major

Source	DF	F	PR > F
Undergraduate Major	13	1.77	0.04

In each case, significant differences were found to exist based on the mean of the residuals by major and post hoc analyses were conducted using Tukey's HSD to determine which means were significantly different.

Discussion

Several majors caused the actual GRE score to be higher than predicted while several other majors caused the actual GRE score to be lower than predicted, although primarily in the area of the GRE quantitative score. In those cases where the actual score was higher, Astin (1991) would say the institution, through its programs and processes within that major was adding value to the student by increasing the GRE score above what would be expected. In the cases where the actual score was lower than predicted, the alternative would be true and the institution would be viewed as holding the student back from his or her true potential. The difference in this approach from a direct use of the GRE scores is that it makes a statistical attempt to adjust for the entering characteristics of students (by taking into consideration the variables used in the model). As discussed earlier, it is not surprising to find that bright students who do well on the SAT also do well on the GRE. The approach discussed above is a method of factoring out the impact of the programs of the postsecondary experience.

The question remaining, from an assessment perspective, is whether this approach provides information which can be used to improve programs and services of the institution. It would appear doubtful. Although the results allow some majors to boast that they enhance the skills of students as defined by performance on the GRE, it does not provide information that could be useful at the curriculum level. For example, knowing that Major X, as a major at this institution, appears to hold back students from their potential in mathematics, provides nothing which could be used directly to improve the program within Major X. Only a gross approach is suggested and that would be to generally strengthen the mathematics portion of the Major X curriculum. Such an approach may not be feasible with accreditation requirements and the normal time-to-degree expectations of students, parents, and legislators.

References

Alexander, J. M., & Stark, J. S. (1986). Focusing on student academic outcomes. Manuscript in preparation, University of Michigan, School of Education.

Angoff, W. H., & Johnson, E. G. (1988). A study of the differential impact of curriculum on aptitude test scores. Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 388 722)

Astin, A. (1991). Assessment for excellence: the philosophy and practice of assessment and evaluation in higher education. New York: American Council on Education/Macmillan Publishing Company.

Banta, T. W., & Associates (1993). Making a difference: outcomes of a decade of assessment in higher education. San Francisco: Jossey - Bass.

Borden, V. M., H & Banta, T.W. (Eds). (1994). Using performance indicators to guide strategic decision making. New Directions for Institutional Research, 82. San Francisco: Jossey-Bass Publishers.

Jacobi, M., Astin, A., & Ayala, F., Jr. (1987). College student outcomes assessment: a talent development perspective (ASHE-ERIC Higher Education Report No. 7). Washington, DC: Association for the Study of Higher Education.

Nichols, J. O. (1991). A practitioner's handbook for institutional effectiveness and student outcomes assessment implementation. New York: Agathon Press.